NATIONAL INSTITUTE OF TECHNOLGY PATNA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
END Semester Examination Jan - June 2022

B. Tech: CSE -VIth semester — Course Name: Data Mining & Warehousing
Course Code: **CS6403 / 7IT166** — Max. Marks : 60
Maximum Time: 3 Hours

Instructions:

1. *Attempt all questions.*
2. *Assume any suitable data, if necessary.*
3. *Answer all parts of the question at the same place.*
4. *Give diagrams where necessary.*

Q.No	Question	Marks	CO	BL
1. a.	What is data warehousing? Explain in short.	**4**	CO2	L1, L2
b.	A database has five transactions. Suppose *min sup = 3*.	**8**	CO3	L3

Transaction Id	Items Bought
100	f, a, c, d, g, i, m, p
200	a, b, c, f, l, m, o
300	b, f, h, j, o, w
400	b, c, k, s, p
500	a, f, c, e, l, p, m, n

Use the priorities of items given below when items have same frequency count: a>b, m>p, f>c, i>g, h>j, b>m, o>w.
Construct an FP-tree with proper labeling and find all the conditional pattern bases and conditional FP trees.

2. a.	What is Decision Tree? Explain Gain Ratio attribute selection measures with formulas?					**4**	CO4	L1, L2
b.	**Outlook**	**Temperature**	**Humidity**	**Wind**	**Play**	**8**		
	sunny	hot	high	FALSE	NO			
	sunny	hot	high	TRUE	NO			
	overcast	hot	high	FALSE	YES			
	rainy	mild	high	FALSE	YES			
	rainy	cool	normal	FALSE	YES			
	rainy	cool	normal	TRUE	NO			
	overcast	cool	normal	TRUE	YES			
	sunny	mild	high	FALSE	NO			
	sunny	cool	normal	FALSE	YES			
	rainy	mild	normal	FALSE	YES			
	sunny	mild	normal	TRUE	YES			
	overcast	mild	high	TRUE	YES			
	overcast	hot	normal	FALSE	YES			
	rainy	mild	high	TRUE	NO			

	In the given dataset there are *two classes* (YES and NO) and four attributes (Outlook, Temperature, Humidity and Wind). Find the *Information* related to attributes Outlook (***Info***$_{Outlook}$), Temperature (***Info***$_{Temperature}$), Humidity (***Info***$_{Humidity}$) and Wind (***Info***$_{Wind}$) up to two decimal places in the process of finding the root node attribute, to construct the decision tree. (*Utilize the Information Gain attribute selection measure.*)		CO4	L1, L3
3. a.	What is Naïve Bayes classifier? Explain with formula.	**4**	CO4	L1, L2
b.				

Colour	***Type***	***Make***	***Classes: Stolen***
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes
Red	Sports	Imported	Yes
Yellow	SUV	Domestic	Yes
Red	Sports	Imported	Yes
Red	SUV	Domestic	No

	Utilize the Naïve Bayes classifier on the above given dataset to find the class of the below given tuple.	**8**	CO4	L3

Color	Type	Origin	Stolen
Red	SUV	Domestic	?

4.	Suppose there are 8 points A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Create three clusters using k-means algorithm and Euclidean distance. Consider initial seeds (Center of each cluster) are A1, A4 and A7. After one round of iteration find all the points in three new clusters and also calculate the centers for all new clusters.	**12**	CO5	L1
5.	Write short notes on the followings:			
a.	Backpropagation in feed forward neural network	**6**	CO4	L1
b.	Major Clustering Methods.	**6**	CO5	L1

**********ALL THE BEST**********